Automatic Text Detection Form Images using Deep Learning Approach

Mr. Macharam[1], Anil Kumar[2]

Dept of Computer Science, Sri Venkateswara University, Tirupati

***Abstract*—** *The adoption of electronic health records (EHRs) is an important step in the development of modern medicine. However, complete health records are not often available during treatment because of the functional problem of the EHR system or information barriers. The medical laboratory report is one kind of important clinical data, which helps health care professionals with patient assessment, diagnosis, and long-term monitoring. The purpose of our work is making papery medical laboratory reports digitalized for EHR system, which mainly relates to optical character recognition (OCR) techniques, especially text detection and recognition. Though OCR is well-established for certain applications, text detection and recognition still face many challenges, such as the many requirements in different scenes (e.g., texts in street scene for robot navigation and receipts OCR etc). This work focuses on the digitization of documents in the medical scene. The most significant challenge to apply a text detection model to a documental image is that the image usually has a high resolution and many textual objects, while the single textual object occupies a very small region. It requires more memory to store the model's variables and takes more time to train and test the model when processing such a large image. A common operation for this problem is to resize the large image into a small scale. This paper presents a deep-learning based approach for textual information segmentation from images of many laboratory reports, which may help physicians solve the data-sharing problem. For text detection, a concatenation structure is designed to combine the features also we can alert the driver around the road signal. The experimental results demonstrate that the text detection in our approach can improve the accuracy of multi-lingual text recognition.*

I. INTRODUCTION

The adoption of electronic health records (EHRs) is an important step in the development of modern medicine. However, complete health records are not often available during treatment because of the functional problem of the EHR system or information barriers. Text detection and recognition has emerged as an important problem in the past few years. Advancements in the field of computer vision and machine learning as well as increase in the applications based on text detection and recognition have resulted in this trend. Various workshops and conferences like International Conference on Document Analysis and Recognition (ICDAR) are being organized on international level giving further rise to developments in field of text processing from imagery. Text detection and recognition from video captions as well as web pages is also getting attention. Huge work has been done in the field of text detection and extraction from natural scenes imagery. Various optical character recognition techniques are also available. In our system, work is making papery medical laboratory reports digitalized for EHR system, which mainly relates to optical character recognition (OCR) techniques, especially text detection and recognition. Though OCR is well-established for certain applications, text detection and recognition still face many challenges, such as the diversified requirements in different scenes (e.g., texts in street scene for robot navigation and receipts OCR for financial departments) and lower quality or degraded data (e.g., scanned legacy books in Google Books service). This work focuses on the digitization of documents in the medical scene. The most significant challenge to apply a text detection model to a documental image is that the image usually has a high resolution and many textual objects, while the single textual object occupies a very small region. It requires more memory to store the model's variables and takes more time to train and test the model when processing such a large image. In this work, a deep learning approach is presented to detect and recognize texts from a laboratory report image. In this approach, a patch-based strategy and a concatenation structure are proposed to handle the problems mentioned above. Specifically, an input documental image is cropped into patches firstly. Then a detector searches textual objects on each patch and outputs a set of predictions. The predictions from all patches are integrated as the final detection results. The module of text recognition is constructed based on CRNN (Convolutional Recurrent Neural Network) and improved through a concatenation structure. For each detected textual object, the text recognizer outputs a text sequence directly. Because mobile devices have been more popular than before, we evaluate the proposed approach on a dataset with both scanned and phone-captured images. The results demonstrate that the proposed approach can effectively detect and recognize texts from medical laboratory reports.

II. LITERATURE SURVEY

E2E-MLT- An Unconstrained End-To End Method for Multi-Language Scene Text.

M. Buta, Y. Patel, and J. Matas (2018)

In this article, we propose an end-to-end trainable (fully differentiable) method for multi-language scene text localization and recognition. The approach is based on a single fully convolutional network (FCN) with shared layers for both tasks. E2E-MLT is the first published multi-language OCR for scene text. While trained in multi-language setup, E2E-MLT demonstrates competitive performance when compared to other methods trained for English scene text alone. The experiments show that obtaining accurate multi-language multi-script annotations is a challenging problem.

Textboxes++: A Single-Shot Oriented Scene Text Detector.

M. Liao, B. Shi, and X. Bai (2018).

In this article, we present an end-to-end trainable fast scene text detector, named Textboxes++, which detects arbitrary-oriented scene text with both high accuracy and efficiency in a single network forward pass. No post-processing other than efficient non-maximum suppression is involved. We have evaluated the proposed Textboxes++ on four public data sets. In all experiments, TextBoxes++ outperforms competing methods in terms of text localization accuracy and runtime. More specifically, TextBoxes++ achieves an f-measure of 0.817 at 11.6 frames/s for 1024 × 1024 ICDAR 2015 incidental text images and an f-measure of 0.5591 at 19.8 frames/s for 768 × 768 COCO-Text images. Furthermore, combined with a text recognizer, TextBoxes++ significantly outperforms the state-of-the-art approaches for word spotting and end-to-end text recognition tasks on popular benchmarks.

Pixellink: Detecting Scene Text via Instance Segmentation

D. Deng, H. Liu, X. Li (2018)

In this article, the paper demonstrated the PixelLink in novel scene text detection algorithm based on instance segmentation, is proposed. Text instances are first segmented out by linking pixels within the same instance together. Text bounding boxes are then extracted directly from the segmentation result without location regression. Experiments show that, compared with regression-based methods, PixelLink can achieve better or comparable performance on several benchmarks, while requiring many fewer training iterations and less training data.

Textsnake- A Flexible Representation for Detecting Text of Arbitrary Shapes.

S. Long, J. Ruan, W. Zhang (2018).

In this article, we propose a more flexible representation for scene text, termed as extit {TextSnake}, which is able to effectively represent text instances in horizontal, oriented and curved forms. In TextSnake, a text instance is described as a sequence of ordered, overlapping disks centered at symmetric axes, each of which is associated with potentially variable radius and orientation. Such geometry attributes are estimated via a Fully Convolutional Network (FCN) model. In experiments, the text detector based on TextSnake achieves state-of-the-art performance on Total-Text and SCUT-CTW1500, the two newly published benchmarks with special emphasis on curved text in natural images, as well as the widely-used datasets ICDAR 2015 and MSRA-TD500. Specifically, TextSnake outperforms the baseline on Total-Text by more than extit in F-measure.

EAST: An Efficient and Accurate Scene Text Detector.

X. Zhou, C. Yao, H.Wen (2017).

In this work, we analyse a simple yet powerful pipeline that yields fast and accurate text detection in natural scenes. The pipeline directly predicts words or text lines of arbitrary orientations and quadrilateral shapes in full images, eliminating unnecessary intermediate steps (e.g., candidate aggregation and word partitioning), with a single neural network. The simplicity of our pipeline allows concentrating efforts on designing loss functions and neural network architecture. Experiments on standard datasets including ICDAR 2015, COCO-Text and MSRA-TD500 demonstrate that the proposed

algorithm significantly outperforms state-of-the-art methods in terms of both accuracy and efficiency. On the ICDAR 2015 dataset, the proposed algorithm achieves an F-score of 0.7820 at 13.2fps at 720p resolution.

Character-Based Handwritten Text Transcription with Attention Networks.

J. Poulos and R. Valle (2017).

In this article, the paper implemented the task of handwritten text transcription with attentional encoder-decoder networks that are trained on sequences of characters. We experiment on lines of text from a popular handwriting database and compare different attention mechanisms for the decoder. The model trained with softmax attention achieves the lowest test error, outperforming several other RNN-based models. Softmax attention is able to learn a linear alignment between image pixels and target characters whereas the alignment generated by sigmoid attention is linear but much less precise. When no function is used to obtain attention weights, the model performs poorly because it lacks a precise alignment between the source and text output.

Focusing Attention- Towards Accurate Text Recognition In Natural Images.

Z. Cheng, F. Bai, Y. Xu (2017).

In this article, we propose the FAN (Focusing Attention Network) method that employs a focusing attention mechanism to automatically draw back the drifted attention. FAN consists of two major components: an attention network (AN) that is responsible for recognizing character targets as in the existing methods, and a focusing network (FN) that is responsible for adjusting attention by evaluating whether AN pays attention properly on the target areas in the images. Furthermore, different from the existing methods, we adopt a ResNet-based network to enrich deep representations of scene text images. Extensive experiments on various benchmarks, including the IIIT5k, SVT and ICDAR datasets, show that the FAN method substantially outperforms the existing methods.

Scene Text Detection and Recognition: Recent Advances and Future Trends.

Y. Zhu, C. Yao, and X. Bai (2016).

In this article, the paper implemented the three-fold: 1) introduce up-to-date works, 2) identify state-of-the-art algorithms, and 3) predict potential research directions in the future. The rich and precise information embodied in text is very useful in a wide range of vision-based applications, therefore text detection and recognition in natural scenes have become important and active research topics in computer vision and document analysis. Moreover, this paper provides comprehensive links to publicly available resources, including benchmark datasets, source codes, and online demos. In summary, this literature review can serve as a good reference for researchers in the areas of scene text detection and recognition.

Text Detection and Recognition in Imagery: A Survey.

Q. Ye and D. Doermann (2015)

In this article, we are analyzing, compares, and contrasts technical challenges, methods, and the performance of text detection and recognition research in color imagery. It summarizes the fundamental problems and enumerates factors that should be considered when addressing these problems. Existing techniques are categorized as either stepwise or integrated and sub-problems are highlighted including text localization, verification, segmentation and recognition. Special issues associated with the enhancement of degraded text and the processing of video text, multi-oriented, perspectively distorted and multilingual text are also addressed. The categories and sub-categories of text are illustrated, benchmark datasets are enumerated, and the performance of the most representative approaches is compared. This review provides a fundamental comparison and analysis of the remaining problems in the field.

Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees.

W. Huang, Y. Qiao (2014).

In this article, we propose a novel framework to tackle this problem by leveraging the high capability of convolutional neural network (CNN). In contrast to recent methods using a set of low-level heuristic features, the CNN network is capable of learning high-level features to robustly identify text components from text-like outliers (e.g. bikes, windows, or leaves). Our

approach takes advantages of both MSERs and sliding-window based methods. The MSERs operator dramatically reduces the number of windows scanned and enhances detection of the low-quality texts. While the sliding-window with CNN is applied to correctly separate the connections of multiple characters in components. The proposed system achieved strong robustness against a number of extreme text variations and serious real-world problems. It was evaluated on the ICDAR 2011 benchmark dataset, and achieved over 78% in F-measure, which is significantly higher than previous methods.

Problem Definition

The adoption of electronic health records (EHRs) is an important step in the development of modern medicine. However, complete health records are not often available during treatment because of the functional problem of the EHR system or information barriers. The medical laboratory report is one kind of important clinical data, which helps health care professionals with patient assessment, diagnosis, and long-term monitoring. Medical laboratory reports digitalized for EHR system, which mainly relates to optical character recognition (OCR) techniques, especially text detection and recognition. Though OCR is well-established for certain applications, text detection and recognition still face many challenges, such as the diversified requirements in different scenes (e.g., texts in street scene for robot navigation and receipts OCR for financial departments) and lower quality or degraded data (e.g., scanned legacy books in Google Books service). In our Existing method the module of text recognition is constructed based on CRNN (Convolutional Recurrent Neural Network) and improved through a concatenation structure. For each detected textual object, the text recognizer outputs a text sequence directly.

Drawbacks

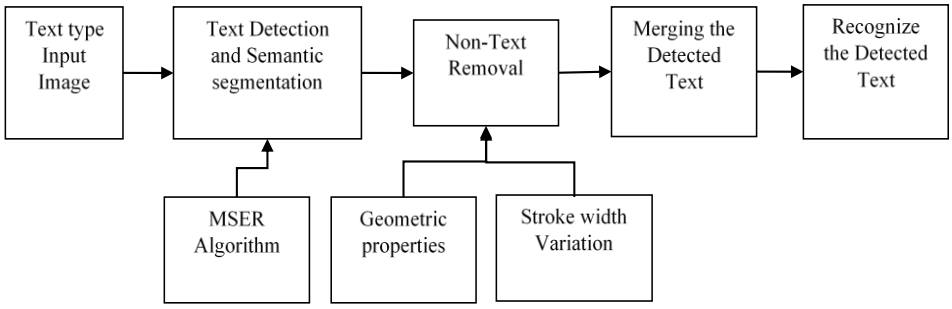
- Not Effectively Detected
- Not Accurate
- Less Performance
- Inefficiency

III. PROPOSED WORK

The adoption of electronic health records (EHRs) is an important step in the development of modern medicine. However, complete health records are not often available during treatment because of the functional problem of the EHR system or information barriers. This paper presents a deep-learning based approach for textual information segmentation from images of many laboratory reports, which may help physicians solve the data-sharing problem. For text detection, a concatenation structure is designed to combine the features also we can alert the driver around the road signal. The experimental results demonstrate that the text detection in our approach can improve the accuracy of multi-lingual text recognition. In our Proposed method to detect regions in an image that contain text. This is a common task performed on unstructured scenes. Unstructured scenes are images that contain undetermined or random scenarios. In real time application for example we can detect and recognize text automatically from captured video to alert a driver about a road sign. This is different than structured scenes, which contain known scenarios where the position of text is known beforehand. Segmenting text from an unstructured scene greatly helps with additional tasks such as optical character recognition (OCR). The automated text detection algorithm detects a large number of text region candidates and progressively removes those less likely to contain text. Finally, the experimental results obtain the better performance when compared to the other system.

Advantages

- Accurate.
- More Text can detect.
- Efficient.
- Better Performance



3.1 Input Image

A collection of data is called dataset. Deep learning requires massive amount of training dataset as classification accuracy of deep learning classifier is largely dependent on the quality and size of the dataset, however, unavailability of dataset is one the biggest barrier in the success of deep learning. Here, an input image can be obtained in the text format image. Text can appear differently in the images.

3.2 Text Detection and Segmentation

The function of text detection is used to determine whether text is present using localization and verification procedures. As a basis of an end-to-end system, it provides precise and compact text instance images for text recognition. Text segmentation has been identified as one of the most challenging problems. It includes text line segmentation and character segmentation. The former refers to splitting a region of multiple text lines into multiple subregions of single text lines. The latter refers to separating a text instance into multiple regions of single characters. Character segmentation was typically used in early text recognition approaches. Here, Semantic Segmentation is used to detect the text from the images.

The automated text detection algorithm is used to detect a large number of text region candidates and progressively removes less likely to contain text. MSER algorithm is used to detect and segment the text from image. The MSER feature detector works well for finding text regions. It works well for text because the consistent color and high contrast of text leads to stable intensity profiles. Use the detect MSER Features function to find all the regions within the image and plot these results. Notice that there are many non-text regions detected alongside the text.

3.3 Nontext Removal

A non-text removal is used to remove the unwanted text from the image. A non-text removal approaches based on the Geometric Properties and stroke width Variation.

3.3.1 Based on the Geometric Properties:

Although the MSER algorithm picks out most of the text, it also detects many other stable regions in the image that are not text. You can use a rule-based approach to remove non-text regions. Geometric properties of text can be used to filter out non-text regions using simple thresholds. Alternatively, you can use a machine learning approach to train a text vs. non-text classifier. Typically, a combination of the two approaches produces better results of the of the system. This example uses a simple rule-based approach to filter non-text regions based on geometric properties.

3.3.2 Based on the stroke width Variation:

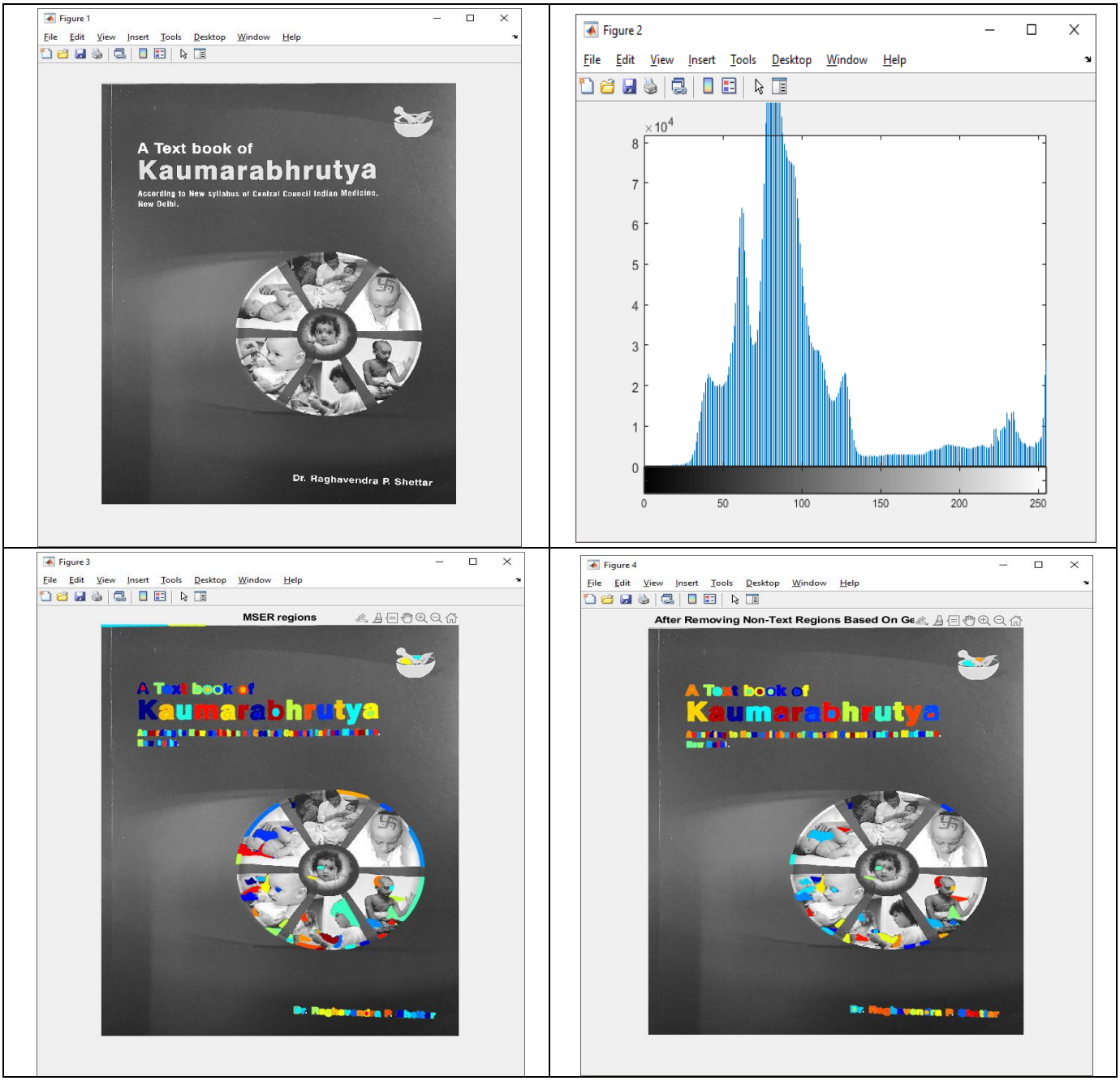
Another common metric used to discriminate between text and non-text is stroke width. Stroke width is a measure of the width of the curves and lines that make up a character. Text regions tend to have little stroke width variation, whereas non-text regions tend to have larger variations. The stroke width can be used to remove non-text regions, estimate the stroke width of one of the detected MSER regions. By using a distance transform and binary thinning operation is used in this model of the system.

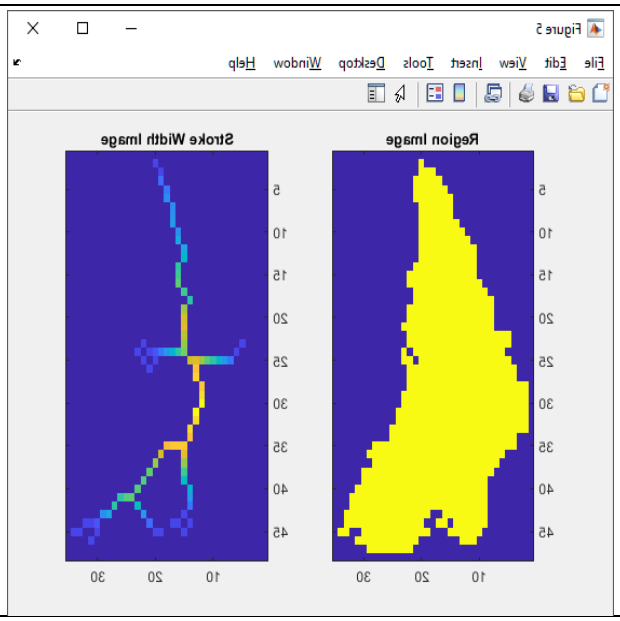
3.4 Merging and Recognizing the Detected Text

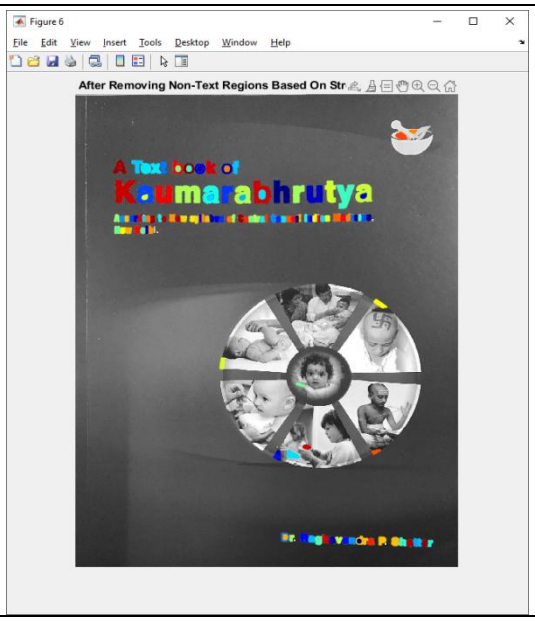
All the detection results are composed of individual text characters. To use these results for recognition tasks, such as OCR, the individual text characters must be merged into words or text lines. This enables recognition of the actual words in an

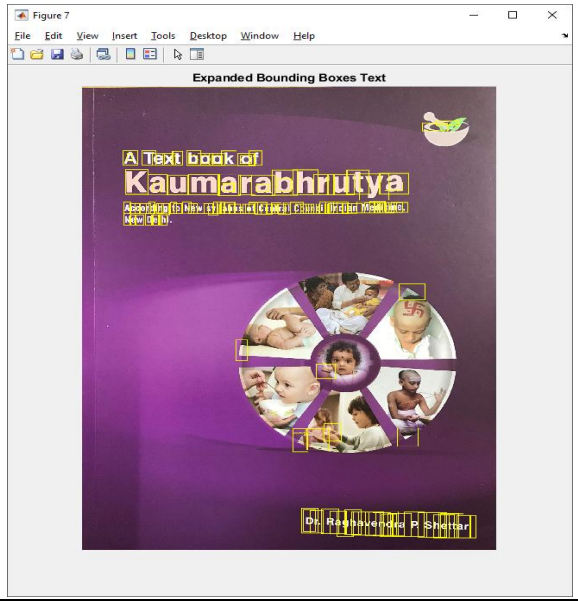
image, which carry more meaningful information than just the individual characters. One approach for merging individual text regions into words or text lines is to first find neighbouring text regions and then form a bounding box around these regions. To find neighbouring regions, expand the bounding boxes computed earlier with region props. This makes the bounding boxes of neighbouring text regions overlap such that text regions that are part of the same word or text line form a chain of overlapping bounding boxes. Now, the overlapping bounding boxes can be merged together to form a single bounding box around individual words or text lines. To do this, compute the overlap ratio between all bounding box pairs. This quantifies the distance between all pairs of text regions so that it is possible to find groups of neighbouring text regions by looking for non-zero overlap ratios. Once the pair-wise overlap ratios are computed, use a graph to find all the text regions "connected" by a non-zero overlap ratio. After detecting the text regions, use the OCR function to recognize the text within each bounding box. The output of the OCR function would be considerably noisier to overcome the noisy region MSER is used to rectify it. Finally, output of the text from the images is predicted.

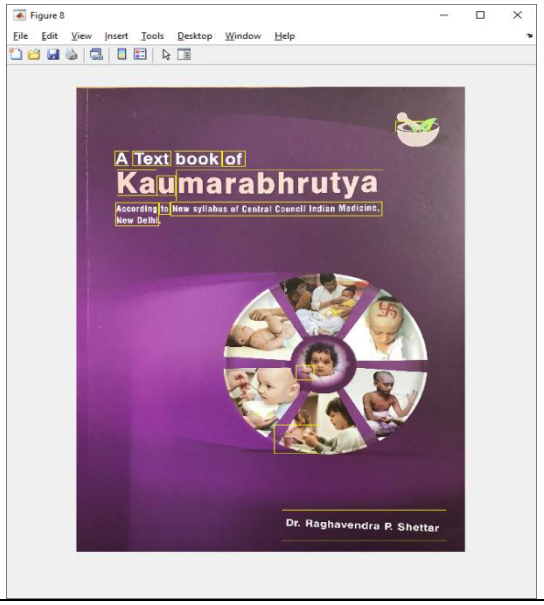
IV. IMPLEMENTATION

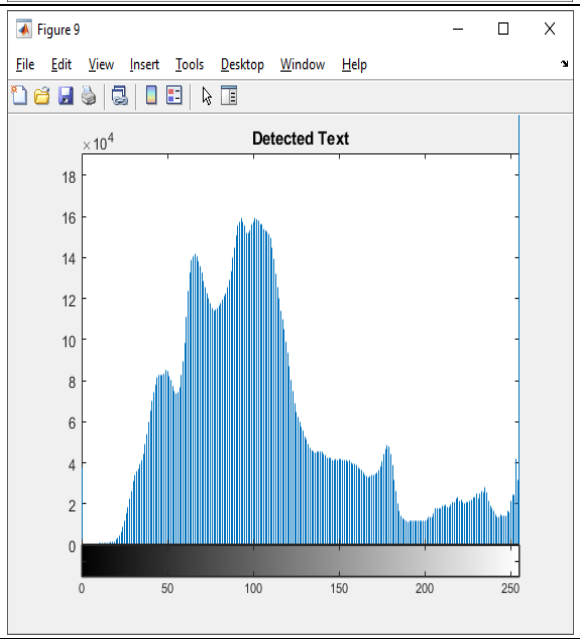












ans =

'Kamarabhrutya

Dr. Raghavendra P. Shettar

book of
According
New Delhi

New syllabus of central council Indian Medicine,

I0

Text

V. CONCLUSION

This paper presents a deep learning approach for text detection and recognition from images of medical laboratory reports. Given an image of medical laboratory report, first, a patch-based training strategy is applied to a detector that outputs a set of bounding boxes containing texts. Then a concatenation structure is inserted into a recognizer, which takes the areas of bounding boxes in source image as inputs and outputs recognized texts. In text detection experiments, image resolution can seriously affect the detection results. Our text detection module is enhanced through a patch-based strategy, which achieves better accuracy, when compared to the other system. The recognition experimental results demonstrate that the concatenation structure can effectively combine shallow and deep features and contribute to the recognition performance. In addition, the experiments on the multi-resolution test set show that the proposed approach has the ability to deal with images with different resolutions. In the End of this paper, we conclude that we can solve the problem of automatic text detection within a natural image is an in many applications. This paper tackles the problem of recognizing text in images captured from the data. The presented approach can be further improved, it would benefit to reducing the cost of manual transcription for digitization of healthcare service in developing countries. The structured health records, which are recovered from document images, will be used for medical data mining to improve health services in our future work.

REFERENCES

[1] Khan, Tauseef & Sarkar, Ram & Mollah, Ayatullah. (2021). Deep learning approaches to scene text detection: a comprehensive review. Artificial Intelligence Review. 54. 1-60. 10.1007/s10462-020-09930-6.

[2] Amritha S Nadarajan, Thamizharasi A, "A Survey on Text Detection in Natural Images", International Journal of Engineering Development and Research (IJEDR), ISSN: 2321-9939, Volume.6, Issue 1, pp.60-66, January 2018.

[3] Tridib Chakraborty et al, (2017), Text recognition using image processing, International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017, 765-768

[4] Karaoglu, S., Tao, R., van Gemert, J. C., & Gevers, T. (2017). Con-Text: Text Detection for Fine-Grained Object Classification. IEEE Transactions on Image Processing, 26(8), 3965-3980.

[5] Guan, L., & Chu, J. (2017, June). Natural scene text detection based on SWT, MSER and candidate classification. In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on (pp. 26-30). IEEE.

[6] Zhu, Q. H., Zhu, R., Li, N., & Yang, Y. B. (2017, October). Deep metric learning for scene text detection. In Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on (pp. 1025-1029). IEEE.

[7] Shi, B., Bai, X., & Belongie, S. (2017). Detecting oriented text in natural images by linking segments. ArXiv preprint arXiv: 1703.06520.

[8] Zhong, Z., Jin, L., & Huang, S. (2017, March). Deeptext: A new approach for text proposal generation and text detection in natural images. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 1208-1212). IEEE.